数理解析研究所講究録
716巻 1990年 124-130

Some Restrictinons on CFGs With Memory

東京女子大学（文理）　守屋悦朗（Etsuro Moriya）

Abstract. Context-free grammars with memory was first introduced by the author in [1]. In this note some restricted versions of them are considered. Among them are (left-/right-/meta-)linear context-free grammars with various types of memory. The memory types considred include stack, pushdown stack, stack counter, counter, queue, and double-ended queue.

§1　諸定義

著者は先に”メモリ付き”文脈自由文法（ＣＦＧＭ）と呼ぶ、文脈自由文法の１つの拡張を定義し、様々のタイプのＣＦＧＭについて考察した。また、それらの特徴付けを分割オートマトンと呼ぶオートマトンによって与えた［１］。ここではそのＣＦＧＭを、基になる文脈自由文法を制限するという観点から考察する。

はじめに、メモリを持つ文脈自由文法（ＣＦＧＭ）の一般的定義を述べる。ＣＦＧＭ (context-free grammar with memory) は、$G = (N, \Sigma, \Gamma, P, S, ¢, \$)$ なる５項組によって指定される。ここに、N は非終端記号の有限集合、Σ は終端記号の有限集合、Γ はメモリ記号の有限集合、$S \in N$ は文記号、$¢, \$ \notin N \cup \Sigma \cup \Gamma$ はメモリのエンドマーカであり、P は次のいずれかの形のプロダクションよりなる有限集合である：

① $A \rightarrow \alpha$　　$(A \in N;\ \alpha \in (N \cup \Sigma)^*)$

② (i) $A \rightarrow Bf$　(ii) $A \rightarrow fB$　　$(A, B \in N;\ f \in \Gamma)$

③ (i) $Af \rightarrow \alpha$　(ii) $fA \rightarrow \alpha$　　$(A \in N;\ \alpha \in (N \cup \Sigma)^*)$

④ $Af \rightarrow Bg$, $Af \rightarrow gB$, $fA \rightarrow gB$, $fA \rightarrow Bg$　　$(A, B \in N;\ f, g \in \Gamma)$

直観的には、ＣＦＧＭとは各非終端記号がメモリを持つＣＦＧのことであり、$¢\Gamma^* N \Gamma^* \$$ の元 $¢\alpha A \beta \$$ は非終端記号Aがメモリ内容 $\alpha\beta$ を持ち、読み書きポインタがαとβの境目のところを指していることを表す。$(¢\Gamma^* N \Gamma^* \$ \cup \Sigma)^*$ の上の２項関係⇒を次のように定義する。$\beta, \gamma \in (¢\Gamma^* N \Gamma^* \$ \cup \Sigma)^*$； $\delta, \varepsilon \in \Gamma^*$； $A_1, \ldots, A_k \in N \cup \Sigma$ に対して、

（１）メモリ内容の継承　$A \rightarrow A_1 \ldots A_k$ が①型プロダクションのとき、

$$\beta ¢ \delta A \varepsilon \$ \gamma \Rightarrow \beta \delta_1 A_1 \varepsilon_1 \ldots \delta_k A_k \varepsilon_k \gamma$$

ここに、$A_i \in N$ のとき $\delta_i = ¢\delta$, $\varepsilon_i = \varepsilon\$$; $A_i \in \Sigma$ のとき $\delta_i = \varepsilon_i = \lambda$

（２）メモリへの書き込み　$A\to Bf$ (or $A\to fB$) が②型のプロダクションのとき、

$$\beta ¢ \delta A \varepsilon \$ \gamma \Rightarrow \beta ¢ \delta Bf \varepsilon \$ \gamma \quad (\text{or } \beta ¢ \delta A \varepsilon \$ \gamma \Rightarrow \beta ¢ \delta fB \varepsilon \$ \gamma)$$

（３）メモリ内容の部分的消去　$Af\to A_1 \ldots A_k$ (or $fA\to A_1 \ldots A_k$)が③型のプロダクションのとき、

$$\beta ¢ \delta Af \varepsilon \$ \gamma \Rightarrow \beta \delta_1 A_1 \varepsilon_1 . \, . \, . \, \delta_k A_k \varepsilon_k \gamma$$

$$(\text{or } \beta ¢ \delta fA \varepsilon \$ \gamma \Rightarrow \beta \delta_1 A_1 \varepsilon_1 . \, . \, . \, \delta_k A_k \varepsilon_k \gamma)$$

ここに、$A_i \in N$ のとき $\delta_i = ¢ \delta$, $\varepsilon_i = \varepsilon \$$; $A_i \in \Sigma$ のとき $\delta_i = \varepsilon_i = \lambda$

（４）メモリポインタの移動　$\zeta \to \theta$ が④型のプロダクションのとき、

$$\beta ¢ \delta \zeta \varepsilon \$ \gamma \Rightarrow \beta ¢ \delta \theta \varepsilon \$ \gamma$$

$\Rightarrow^*$を$\Rightarrow$の反射推移閉包とするとき、

$$L(G) = \{ w \in \Sigma^* \mid ¢ S\$ \Rightarrow^* w \}$$

を、G の生成する言語（<u>ＣＦＬＭ</u>）という。

さて、［１］ではＣＦＧＭの持つメモリ構造にいくつかの制限を加えたものを考えた。

（ａ）④型のプロダクションを含まないとき、<u>ＣＦＧ２Ｐ</u> (context-free grammar with two pushdown stores)という。

（ｂ）ＣＦＧ２Ｐで、かつ、$\Gamma = \{¥, Z\}$、かつ、任意の文形式が $(¢ ¥Z^*NZ^*¥\$ \cup ¢ NZ^*¥\$ \cup ¢ ¥Z^*N\$ \cup ¢ N\$ \cup \Sigma)^*$ の形をしているようなＣＦＧＭを<u>ＣＦＧ２Ｃ</u> (context-free grammar with two counters) という。¥は、カウンタの底を表す。

（ｃ）次に、②(ii)、③(ii)型のプロダクションを許さず、かつ、④型のプロダクションは次の形のものだけに制限する：

$$Af\to Bf, \; Af\to fB, \; fA\to fB, \; fA\to Bf$$

さらに、②(i)型プロダクション $A\to Bf$、③(i)型プロダクション $Af\to \alpha$ が適用出来るのは、それぞれ次の形の文形式に限定する：

$$\beta ¢ A \delta \$ \gamma \quad \text{または} \quad \beta ¢ Af \delta \$ \gamma$$

$$(A, B \in N;\; f \in \Gamma;\; \delta \in \Gamma^*;\; \beta, \gamma \in (¢ \Gamma^* N \Gamma^* \$ \cup \Sigma)^*)$$

このようなＣＦＧＭを <u>ＣＦＧＳ</u> (context-free grammar with a stack) という。

（ｄ）ＣＦＧＳで、かつ、③型のプロダクションを含まないものを<u>ＣＦＧＮＥＳ</u> (context-free grammar with a nonerasing stack) という。

（ｅ）ＣＦＧＳで、かつ、$\Gamma = \{¥, Z\}$、かつ、任意の文形式が $(¢ NZ^*¥\$ \cup ¢ N\$ \cup \Sigma)^*$ の形をしているようなものを<u>ＣＦＧＳＣ</u> (context-free grammar with a stack counter) という。

（ｆ）④型のプロダクションを含まないＣＦＧＳを<u>ＣＦＧＰ</u> (context-free grammar with a pushdown store) という。これは、インデックス文法［３］に他ならない。

（ｇ）ＣＦＧＰで、かつ、$\Gamma = \{¥, Z\}$、かつ、任意の文形式が $(¢ NZ^*¥\$ \cup ¢ N\$ \cup \Sigma)^*$

の形をしているようなものをCFGC (context-free grammar with a counter)という。

X型の文法によって生成される言語のクラスを $\mathcal{L}_X$ で表すことにする。［1］［2］で次のことが示されている。

定理1（1）$\mathcal{L}$(CFG2C)＝$\mathcal{L}$(CFG2P)＝$\mathcal{L}$(CFGM)＝帰納的可算言語のクラス
（2）$\mathcal{L}(\mathrm{CFG}) \subsetneq \mathcal{L}(\mathrm{CFGC}) \subsetneq \mathcal{L}(\mathrm{CFGP}) \subsetneq \mathcal{L}(\mathrm{CFGS}) \subset \mathcal{L}(\mathrm{CSG})$

§2　制限付きCFGM

ここでは、CFGMの基礎になるCFGに制限を加えることを考える。$G = (N, \Sigma, \Gamma, P, S, ¢, \$)$ をCFGMとする。$h:(N \cup \Sigma \cup \Gamma)^* \to (N \cup \Sigma)^*$ を、$A \in N \cup \Sigma$ のとき $h(A) = A$、$f \in \Gamma$ のとき $h(f) = \lambda$ で定義された準同型写像とし、

$$\underline{P} = \{h(x) \to h(y) \mid x \to y \in P\}$$

と定義する。$\underline{G} = (N, \Sigma, \underline{P}, S)$ を G の基礎CFG (underlying context-free grammar) という。$\underline{G}$ が右線形（左線形、線形）文法であるとき、G は右線形（左線形、線形）CFGMであるという。

Aho［3］は、CFGSにおいて①型のプロダクションと③型のプロダクションのタイプ分けを行った。①がα型、かつ、③がβ型であるCFGSを$\alpha\beta$型文法ということにする。従って、右線形CFGP＝右線形右線形CFGPである。次の結果が既知である［2］。

定理2（1）$\mathcal{L}$(CFG)＝$\mathcal{L}$(右線形右線形CFGP)＝$\mathcal{L}$(左線形左線形CFGP)
（2）$\mathcal{L}$(CFG) $\subsetneq$ $\mathcal{L}$(右(左)線形左(右)線形CFGP)

実際、定理2（1）の左側の等式はもっと一般に成り立つ。すなわち、メモリタイプαの右線形右線形CFGSはメモリタイプαの非決定性1方向オートマトンと能力が等しい。以下では、非決定性1方向スタックオートマトン［5］を”SA”、非決定性1方向スタック非消去型スタックオートマトン［5］を”NESA”、非決定性1方向1スタックカウンタマシン［4］を”SCA”、非決定性1方向1カウンタマシン［6］を”CA”と略記し、これらが受理する言語のクラスをそれぞれ$\mathcal{L}$(SA), $\mathcal{L}$(NESA), $\mathcal{L}$(SCA), $\mathcal{L}$(CA)と書く。

命題1　$X \in \{S, NES, SC, C\}$に対して、$\mathcal{L}$(右線形右線形CFGX)＝$\mathcal{L}$(XA).

証明　X=S の場合について考える。$\mathcal{L}$(SA)$\subset$$\mathcal{L}$(右線形右線形CFGS)の証明は、SAの状

態をＣＦＧＳの非終端記号とし、入力記号の読み込みと状態変化を右線形プロダクションでシミュレートする。ＳＡのスタック読みモード（状態 A でスタック内容 $\beta\gamma$ のβとγの境目にポインタがあり、入力記号 a を読んでいるとする)に対応して、ＣＦＧＳではスタック内にある非終端記号（$¢\beta A\gamma\$$ の A が対応する）に①型の右線形プロダクション $A \rightarrow a\underline{A}$ とそれらに引き続いてスタックポインタの移動に対応する④型のプロダクションを適用すればよい。

$\mathcal{L}$(右線形右線形CFGS)⊂$\mathcal{L}$(SA)の証明も同様であるが、ＣＦＧＳにおけるスタック読みモード（④型プロダクションの適用）に対しては、ＳＡにおいてはλ-入力動作が対応する。

X∈{NES，SC，C}の場合も全く同様である。

どのプロダクションの右辺にも終端記号が含まれていないようなＣＦＧＳを実時間ＣＦＧＳという。realtime ＣＡが ^{R} で閉じていない［６］ことから次を得る。

系　$\mathcal{L}$(実時間右線形右線形CFGS)は ^{R} で閉じていない。

補題１　X∈{CFGS，CFGNES，CFGSC，CFGC}に対して、

$L \in \mathcal{L}$(右線形右線形X) ⇔ $L^R \in \mathcal{L}$(左線形左線形X),

$L \in \mathcal{L}$(右線形左線形X) ⇔ $L^R \in \mathcal{L}$(左線形右線形X).

証明　容易。

命題２　X∈{CFGS，CFGP}に対して、$\mathcal{L}$(右線形右線形X)＝$\mathcal{L}$(左線形左線形X).

証明　$L \in \mathcal{L}$(右線形右線形X) ⇔ $L^R \in \mathcal{L}$(左線形左線形X)

⇔ $L \in \mathcal{L}$(左線形左線形X)

前半の⇔は補題１より、後半の⇔は ＣＦＬ，ＳＡ が ^{R} で閉じていることによる。

その他のクラスについても、^{R} で閉じていることが言えれば同様のことが成り立つ。命題２の証明から、

$\mathcal{L}$(左線形右線形X)が ^{R} で閉 ⇔ $\mathcal{L}$(左線形右線形X)＝$\mathcal{L}$(右線形左線形X)

⇔ $\mathcal{L}$(右線形左線形X)が ^{R} で閉

が成り立つことに注意したい。

今のところ、X∈{S，A}について、$\mathcal{L}$(XA)と$\mathcal{L}$(左線形右線形CFGX)、$\mathcal{L}$(右線形左線形CFGX)との包含関係については不明であるが、次のことは容易に示すことが出来る。正則集合のクラスを$\mathcal{R}$で表す。

補題２　$L_0 = \{a^{n \cdot n} \mid n \geqq 0\} \in \mathcal{L}$(CFGP).

<u>証明</u> 次のプロダクションより成るＣＦＧＰ $G=(\{S',S,A,A',B\},\{a\},\{f,g\},P,S,¢,\$)$ を考える。

$$S'\to Sg$$
$$S\to AB \mid \lambda$$
$$A\to A'f,\qquad A'\to Sf$$
$$Bf\to aB,\qquad Bg\to a$$

明らかに

$$S'\Rightarrow^{*}Af^{2n-2}gBf^{2n-2}gBf^{2n-4}g\ldots Bf^{2}gBg$$
$$\Rightarrow^{*}Sf^{2n}gBf^{2n-2}gBf^{2n-4}gg\ldots Bf^{2}gBg \qquad (n=0,1,2,\ldots)$$
$$\Rightarrow Bf^{2n-2}gBf^{2n-4}g\ldots Bf^{2}gBg$$
$$\Rightarrow^{*}a^{2n-1}a^{2n-3}\ldots a^{3}a$$

で、$\Sigma_i(2i-1)=n^2$ であることに注意すると $L(G)=L_0$ である。

<u>命題３</u>（１）（ａ）<u>$\mathcal{R}\subsetneq\mathcal{L}$(左線形右線形CFGC)$\cap\mathcal{L}$(右線形左線形CFGC)</u>

（ｂ）<u>$\mathcal{L}$(左線形右線形CFGC)$\cup\mathcal{L}$(右線形左線形CFGC)$\subsetneq\mathcal{L}$(CFGP)</u>

（ｃ）<u>$\mathcal{L}$(左線形右線形CFGC)$-\mathcal{L}$(CA)$\neq\phi$</u>

<u>$\mathcal{L}$(右線形左線形CFGC)$-\mathcal{L}$(CA)$\neq\phi$</u>

（２）<u>$\mathcal{L}$(CFG)$\subsetneq\mathcal{L}$(左線形右線形CFGP)$\subsetneq\mathcal{L}$(CFGP)</u>

<u>$\mathcal{L}$(CFG)$\subsetneq\mathcal{L}$(右線形左線形CFGP)$\subsetneq\mathcal{L}$(CFGP)</u>

<u>証明</u>（１）（ａ）右（左）線形文法をシミュレートすればよいので$\subset$は明らか。また、$\{a^nb^n \mid n\geqq 0\}\in\mathcal{L}$(左線形右線形CFGC)$\cap\mathcal{L}$(右線形左線形CFGC)$-\mathcal{R}$ は容易に示せる。

（ｂ）$\subset$は自明。ΨをParikh写像とするとき、$L\in\mathcal{L}$(左線形右線形CFGC)$\cup\mathcal{L}$(右線形左線形CFGC) ならば $\Psi(L)$ は準線形集合(semilinear set)である。なぜなら、例えば $L\in\mathcal{L}$(左線形右線形CFGC) とするとき、$L=L(G)$ となる左線形右線形ＣＦＧＣに対して、G の③型のプロダクション $Af\to Ba$ を $Af\to aB$ に置き換えた文法 G' を考えると G' は右線形右線形ＣＦＧＣであり、しかも $\Psi(L(G))=\Psi(L(G'))$ である。命題１より $L(G')\in\mathcal{L}$(CA)$\subset\mathcal{L}$(CFG) であるから、$\Psi(L(G'))$ は準線形集合である。

よって、補題２より$\subsetneq$である。

（ｃ） $L=\{a^ib^ja^ib^j \mid i,j\geqq 1\}$ は次のプロダクションを持つ左線形左線形ＣＦＧＣによって生成される。S は文記号、A, A', B, B', C は非終端記号、f, g, h はメモリ記号である。

$$S\to Ah$$
$$A\to aA' \mid B,\qquad A'\to Af$$
$$B\to bB' \mid C,\qquad B'\to Bg$$
$$Cf\to Ca,\qquad Cg\to Cb,\qquad Ch\to\lambda$$

補題１より $L^R\in\mathcal{L}$(右線形左線形CFGC) であり、L, L^R ともにＣＦＬでないから$\mathcal{L}$(CA)の

元ではない。

（２）$\mathcal{L}$(CFG)$\subsetneq\mathcal{L}$(左線形右線形CFGP)、$\mathcal{L}$(CFG)$\subsetneq\mathcal{L}$(右線形左線形CFGP) は［３］で示されている。$\mathcal{L}$(左線形右線形CFGP)$\subset\mathcal{L}$(CFGP)、$\mathcal{L}$(右線形左線形CFGP)$\subset\mathcal{L}$(CFGP) は自明。これらの包含関係が$\subsetneq$であることは（１ｂ）と同様に示される。命題２より $\mathcal{L}$(右線形右線形CFGP)=$\mathcal{L}$(CFG) であることに注意したい。

$\mathcal{L}$(左線形右線形CFGS)$\cup\mathcal{L}$(右線形左線形CFGS)$\subsetneq\mathcal{L}$(CFGS) か否かは未解決である。
次の命題は命題３のanalogyである(証明方法も)。

<u>命題４</u>　<u>$\mathcal{L}$(CFG)$\subsetneq\mathcal{L}$(線形CFGP)$\subsetneq\mathcal{L}$(CFGP).</u>

ＣＦＧ G = (N, Σ, P, S) において、どのプロダクションも線形であるか、または、$S\rightarrow\alpha$ （$\alpha\in(N\cup\Sigma)^*$）の形をしており、かつ、どのプロダクションの右辺にも S が現われないとき、G は<u>超線形</u>［８］であるという。補題２の L_0 が超線形ＣＦＧＰで生成されていることより、次を得る。

<u>命題５</u>　<u>$\mathcal{L}$(線形CFGP)$\subsetneq\mathcal{L}$(超線形CFGP)</u>

$\mathcal{L}$(超線形CFGP)$\subsetneq\mathcal{L}$(CFGP)か、$\mathcal{L}$(超線形CFGP)$\subsetneq\mathcal{L}$(超線形CFGS)か否かなどは未解決である。

メモリ構造としては、first-in last-out方式の（プッシュダウン）スタックに対応するものとしてfirst-in first-out方式のキューが考えられる。また、キューの内部を読むことの出来る（書き換えは不可）ようなキューの拡張も考えられる。スタックの場合と異なり、キューの場合にはキューが空であるか否かを判定する能力を持たせるか否かで能力に差が生じるものと思われる。ここでは、空であるか否かを判定する能力を持たないモデルを考える（文法による定義ではこの方が自然である）。キューをメモリ構造として持つＣＦＧＭを<u>ＣＦＧＱ</u> (context-free grammar with a queue)という。また、double-ended queueをメモリ構造として持つＣＦＧＭを<u>ＣＦＧＤＥＱ</u>という。

<u>命題６</u>　<u>X∈{左線形左線形，左線形右線形，右線形左線形，右線形右線形}について、$\mathcal{L}$(XCFGC)$\subset\mathcal{L}$(XCFGQ)．特に、$\mathcal{L}$(CFG)$\subsetneq\mathcal{L}$(CFGQ).</u>

<u>証明</u>　ＸＣＦＧＱにおいてキューが空か否かを判定出来れば$\subset$が成り立つ。ＸＣＦＧＣが初めにカウンタの底にbottom markerを書き込むのに対応して、ＸＣＦＧＱは非決定性にある時点でend-of-queue markerを書き込んでやることにより、このシミュレーションが可能である。

double-ended queueは

fff..ff#ggg...gg

の形で２本のカウンタをシミュレート出来る（＃は両方のカウンタのbottom markerの役割を果たす）ので、［２］の $\mathcal{L}$(CFG2C)＝帰納的可算言語のクラス の証明より次を得る。

命題７ $\mathcal{L}$(CFGDEQ)＝帰納的可算言語のクラス.

Ａｈｏ［３］は、右（左）線形右（左）線形以外にもいくつかの制限付ＣＦＧによるインデックス文法（＝ＣＦＧＰ）を考えて興味深い結果を示している。ＣＦＧＭにおいて同様なことを考えてみるのも価値があろう。

謝辞　辛抱強く、この小考察を書く機会を下さった、研究代表者の一橋大学の町田元氏に感謝します。

文献

1. 守屋悦朗、ある拡張文法／オートマトンに関するコメント、数理解析研究所講究録695 (1989)、21-26.
2. E.Moriya, Context-free grammars with memory, submitted for publication, 1989.
3. A.V.Aho, Indexed grammars - an extension of context-free grammars, J.ACM 15 (1968), 647-671.
4. R.V.Book and S.Ginsburg, Multi-stack-counter languages, Math. Systems Theory 6 (1972), 37-48.
5. S.Ginsburg, S.A.Greibach and M.A.Harrison, One-way stack automata, J.ACM 14 (1967), 389-418.
6. P.C.Fischer, A.R.Meyer and A.L.Rosenberg, Counter machines and counter languages, Math. Systems Theory 2 (1968), 265-283.
7. J.E.Hopcroft and J.D.Ullman, Nonerasing stack automata, JCSS 1 (1967), 166-186.
8. S.Ginsburg, "The Mathematical Theory of Context-Free LAnguages," McGraw-Hill, N.Y., 1966.